Bandolier 55

What do we think?
What do we know?
What can we prove?

Evidence-based health care £3.00 September 1998 Volume 5 Issue 9

Spinning wheels

This month ***Bandolier*** picks out a clever systematic review of audits of the inappropriate use of clinical laboratory tests. It shows that an awfully large proportion of tests are inappropriate. This makes for a lot of wheel spin in an area which consumes a significant chunk of NHS spending. We do some back-of-an-envelope calculations to show what this means in financial terms.

This is not just about potentially saving money. There is also an opportunity cost. The huge pressures of ever increasing workloads for laboratories often prevents the introduction of better, though perhaps more expensive, tests. Nor do studies often tell us about tests which should have been done which weren't.

Bandolier would like to tell more positive stories about tests, about how test results might be better be presented, and about how tests may be employed to reduce health care costs. They seem awfully hard to find. We would like your help, therefore, in locating papers which report how tests help in diagnosis or prognosis, or help in defining or delivering a service, or in helping provide a higher quality or cheaper service.

Getting a grip

The number-needed-to-treat (NNT) has provided a common currency for describing the intervention-specific benefits that accrue from treatments. A report from Wessex shows that about 35% of GPs are able to understand NNTs and feel confident about explaining it to others. Given that a few years ago none of us knew anything about NNTs, that's quite an advance, probably because NNTs make sense and are understandable.

But new readers and thinkers join all the time, so this month we reprise the calculation of NNTs and NNHs. We also look at where GPs get their evidence-based information. OK, so ***Bandolier*** looks good in that report on page 6, but the main point is that awareness of evidence-based information is low.

Getting *Bandolier*

Which brings us to the business of getting ***Bandolier***. This is given in detail again, but of course much will change with the new Regions next year. Many people photocopy ***Bandolier*** we hear. That may be more wheel spin. It is probably more expensive than asking for more copies. The extra cost of producing and distributing more copies of ***Bandolier*** is negligible - tripling the distribution would add less than 10% to the total cost. So if you want copies for each partner, nurse, or manager, or for community pharmacists, then don't fret or copy, just ask and we'll do our best to help.

In this issue

THE SAME AS YESTERDAY, LADDIE!

We use laboratories with ever greater frequency to order tests to help make or exclude diagnosis, or help develop a prognosis, or because we always order these tests in these sorts of patients. Sometimes the ordering of those tests is inappropriate.

Bandolier well remembers an occasion, many years ago, when a new house officer darted into a laboratory on a Friday evening at 5.00 pm demanding an "urgent" albumin measurement on a patient. A wise old biochemist (one of ***Bandolier***'s heroes) asked whether a test had been done previously.

Yes, one had been done yesterday.
What was the result then?
42 (or whatever albumin results are).
So the sample was held up to the sunset, and the wise old biochemist said "It's the same today, laddie, the same today".

The point is that albumin concentrations don't really change that much unless something major has happened, and it hadn't in this case. The requesting of an urgent albumin in this case was inappropriate.

Review

David Naylor and Carl van Walraven have done a heroic job in trying to discover how best to define inappropriateness, and to find out how big the problem of inappropriate laboratory testing is by systematically reviewing laboratory audits [1]. Their paper is long and complicated, but full of interesting thoughts, most of which ***Bandolier*** can only skate over.

Their searching was thorough, and they restricted themselves by excluding radiological or pathological tests, and screening tests. They found 44 articles, 34 of which had explicit criteria for appropriateness (and which looked generally to ***Bandolier*** to be fair, or even conservative).

Summary of inappropriate test use

Study	Number of reports	Number of tests	Percent inappropriate	Range
Studies with implicit criteria	11	5360	56	11 - 95
General biochemistry & haematology	5	63,030	15	11 - 70
Microbiology	7	4979	46	5 - 95
Cardiac enzymes	2	843	39	38 - 96*
Thyroid function	4	2490	30	17 - 55
Drug monitoring	16	2787	46	5 - 83

*range includes data not used for calculation of overall mean

Results

The main results from this enormous amount of work are summarised in the Table, where information has been pooled to give an estimate of the overall inappropriate use for different categories. The clear answer is that tests are used inappropriately to a very large extent, which in some cases is over 90%. We do not know the corollary - that is, how often should tests be used but are not, since few studies examined that aspect.

Comment

Anyone with a personal or professional interest in this needs to read the paper in its entirety, probably at least three times. The authors pack the discussion with good sense, including a paradigm for the classification of test appropriateness.

Clearly more work needs to be done. Radiology, pathology and screening tests were omitted. Results there may be different, because radiologists have worked hard to ensure the appropriateness of tests which may cause harm, as exemplified by the Ottawa ankle and knee tests in particular (see ***Bandolier*** 12, 21, 49). The reports almost all emanated from sick patients in teaching hospitals. What difference would there be for out-patients, or primary care, or non-teaching hospitals? So the best guess at the moment is that inappropriate use of laboratory tests is uncomfortably common, something that most laboratory personnel would recognise.

Financial implications

The implications for costs are significant, even though each intervention is in itself small - an example of the technology creep identified in the very first issue of ***Bandolier***. In the UK we spend £40 billion or so on the NHS, and an estimate of costs of laboratories and laboratory testing is 4% of that, or £1.6 billion. Then it's a question of choosing a figure for inappropriate tests. If it is only 10%, that is £160 million wasted, but the figures in the Table suggest that may be conservative. The real answer is probably unknowable, but whatever it is, it is large.

And the UK has one of the best, and cheapest, laboratory services. In Europe and elsewhere laboratory costs are higher as a percentage of total healthcare costs, rising to perhaps twice or more in the USA.

Where do we go from here?

George Lundberg, editor of JAMA, pens an accompanying editorial [2] which emphasises the need for an outcomes research agenda. While he is right, the hope may be a pious one, because however important the issues raised by the review, the fact is that this is not seen as an important issue, neither academically in our Universities, nor managerially in the NHS. We start from a low base and there are no resources to do this work. Lundberg begins his editorial with words he used in 1975 which encapsulate the problem. Lets hope the same words don't have to be repeated in 2021.

References:
1 C van Walraven, CD Naylor. Do we know what inappropriate laboratory utlization is? A systematic review of laboratory clinical audits. JAMA 1998 280: 550-8.
2 GD Lundberg. The need for an outcomes research agenda for clinical laboratory testing. JAMA 1998 280: 565-6.

EDITORS

Dr Andrew Moore
Dr Henry McQuay Dr J A Muir Gray

Pain Relief Unit
The Churchill, Oxford OX3 7LJ

Editorial office: 01865 226132
Editorial fax: 01865 226978
Email: andrew.moore@pru.ox.ac.uk
Internet: http://www.jr2.ox.ac.uk/Bandolier

ISSN 1353-9906

Acid commentary on pH

One often hears the comment "If it ain't broke, why fix it?". A thoughtful and beautifully written piece [1] on the lack of understanding of pH tells us this is broke, and that it needs fixing. Try and answer these simple questions:

Question 1

If the pH of a solution is 7.40, what is the new pH:

(a) if the hydrogen ion concentration doubles? (hint - the logarithm of 2 is 0.3).
(b) if the hydrogen ion concentration increases 10-fold?

Question 2

If the hydrogen ion concentration of a solution is 40 nmol/L, what is the new hydrogen ion concentration:

(a) if the hydrogen ion concentration doubles?
(b) if the hydrogen ion concentration increases 10-fold?

The answers to the first question are 7.1 and 6.4. If you can't work out the answers to the second, then you are beyond help. For you to check it out, the Table and Figure give the relationship between hydrogen ion concentration in nmol/L and pH.

Clever medical students

When these questions were put to 430 medical students about to take their final pathology examinations, only 10% could get the answers to question 1 correct. Now don't forget that these are medical students who have, nowadays, to get virtually all A grades at A level, usually including mathematics, and that to get into medical school a Nobel prize or two helps. So if only 10% of them can get it right, what hope is there for the non-anaesthetists among us. The students found the answers to the second question trivial.

Surveys

Most respondents to an external quality scheme (148/157) report pH rather than hydrogen ion concentration, though rather fewer (72%) of 22 medical schools do so. So pH lives on in most UK hospitals even if understood by few new doctors (and maybe even older ones?).

Veil of mystery

The point the paper makes, as others have made before, is that pH is an inverse logarithm, and is counter-intuitive. It is hopeless trying to make corrections against changes in carbon dioxide concentrations without complex algorithms. A nice turn of phrase likens pH to a "veil of mystery diverting attention from the leading character within the drama of acid-base homeostasis, namely the hydrogen ion concentration".

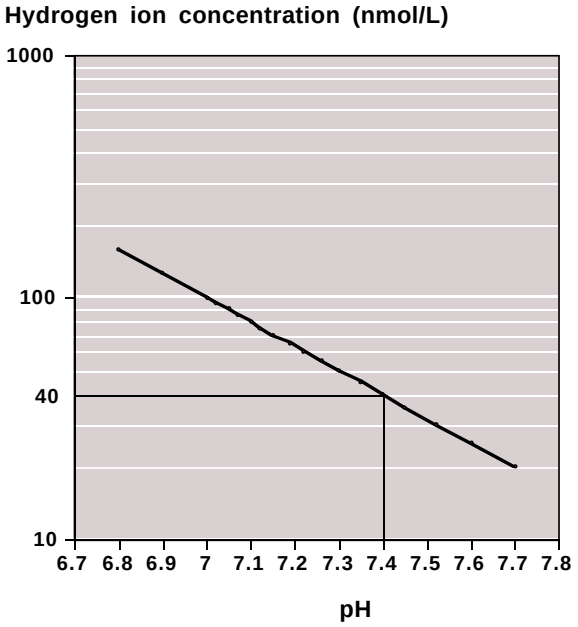
Comment

It is not as if this is a new finding, and the authors show how the consequences of poor understanding of pH, and arguments about measurement and description of pH and hydrogen ion concentration have been rehearsed before. But nothing has been done. ***Bandolier*** got the correct answers to question 1, but gets a severe headache when asked to think about acid-base homeostasis. It has become a subject best avoided. If that is because we can't understand pH, and there is a better and more easily understood method using hydrogen ion concentrations in nmol/L, then it's time for a change. There's no point in having evidence if people can't use it. We've dumped odds ratios in favour of NNTs, so why not dump pH in favour of nmol/L?

Reference:
1 J Hooper, WJ Marshall, AL Miller. Log-jam in acid-base education and investigation: why make it so difficult? Annals of Clinical Biochemistry 1998 35: 85-93.

Relationship between hydrogen ion concentration and pH

Hydrogen ion concentration (nmol/L)	pH
20	7.70
25	7.60
30	7.52
35	7.45
40	7.40
45	7.35
50	7.30
55	7.26
60	7.22
65	7.19
70	7.15
75	7.12
80	7.10
85	7.07
90	7.05
95	7.02
100	7.00
126	6.90
158	6.80



TONSILLECTOMY FOR SORE THROATS

Tonsillectomy is a common operation, with over 70,000 being carried out every year. The main reason for doing it is to prevent recurrent throat infection - either tonsillitis or pharyngitis. How well does tonsillectomy work?

Review

A systematic review [1] searched widely and obtained some unpublished information. The entry criteria were participants suffering from recurrent throat infections, the operation was tonsillectomy or adeno-tonsillectomy, and studies needed to be randomised.

Results

Five studies were found, two of which had strict entry criteria related to frequency of throat infections and defining features of a throat infection. The other three had less explicit criteria. No trial included children over 15 years.

The main outcome was that of incidence of throat infection during two years of follow up. For all throat infections the results for years 1 and 2 are seen in the Figure. Over the two years, tonsillectomy meant avoiding 2.3 to 3.6 throat infections. Of these infections, perhaps 10-30% would be likely to be moderate or severe. More infections occurred in control groups in the first than in the second year. One study had a three year follow up, by which time the benefits of tonsillectomy had been lost. There was no statistical significance in school days lost between treated children and controls.

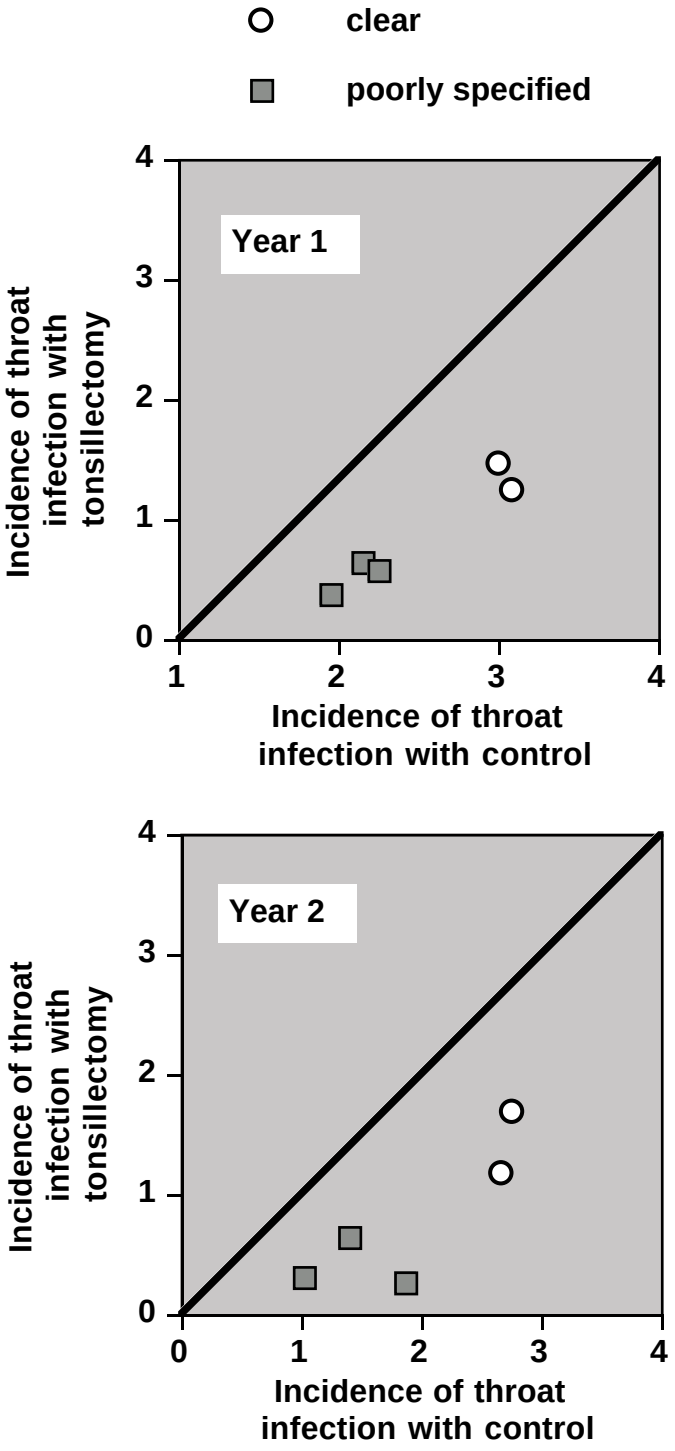
Comment

Four of these trials were conducted over 10 years ago, and three about 30 years ago, though that need not diminish their value. Tonsillectomy is itself not without adverse effects. These include post-operative bleeding (though new laser techniques may avoid that), as well vomiting and fever.

Any decision on tonsillectomy should not be taken lightly, and must be a balance between possible benefits and harms. Maybe there is work being done to define the benefits and harms more clearly than in the past. We'd love to see it.

Reference:
1 T Marshall. A review of tonsillectomy for recurrent throat infection. British Journal of General Practice 1998 48: 1131-5.

Effect of tonsillectomy on subsequent throat infections in the two years after operation: studies in which entry criteria were



TURNING EVIDENCE INTO EVERYDAY PRACTICE

A series of seminars is being held around the UK to let people share the experiences of the King's Fund PACE projects. The mornings will be devoted to presentations and the afternoons to panel discussions and small group work. The seminars will explain how change has been managed to implement best practice and best evidence. Details are available from Hayley Hutton on 0171 307 2694. Seminars will be held at:

Reading: Tuesday October 13th
Warrington: Monday November 16th
Leeds: Monday November 2nd
London: Wednesday December 2nd

OVARIAN CANCER RISK

With ovarian cancer (overall incidence for women aged 45 years is about 1 in 10,000) there is an element of extra risk emanating from genetic associations, for instance the BRCA1 and BRCA2 genes (see ***Bandolier*** 18 & 44). A new meta-analysis from Cambridge pools information to obtain estimates of the increased risk women may have from having relatives affected with ovarian cancer.

Searching and studies

MEDLINE and hand searching found 13 case-control studies and three cohort studies. One of the cohort studies was omitted because it was thought to be biased. Information was pooled where appropriate to obtain relative risks, and estimates were calculated of the lifetime risk to relatives up to age 75 years.

Results

The main results are shown in the Figure. In detail, the results for each circumstance were as follows:

Second degree relative

The pooled relative risk from three studies for having an affected second degree relative was 2.5 (95% CI 1.5 to 4.3).

First degree relative

The pooled relative risk from 15 studies for having an affected first degree relative was 3.1 (95% CI 2.6 to 3.7). The risk of developing ovarian cancer by age 75 is about 4% for women younger than 45 with an affected first degree relative. The risk falls off quite rapidly after this age.

Affected sister

The pooled relative risk from four studies for having an affected sister was 3.8 (95% CI 2.9 to 5.1). The risk of developing ovarian cancer by age 75 is about 5% for women younger than 45 with an affected sister. The risk declines after this age, but remains substantially higher than the general population.

Affected mother

The pooled relative risk from three studies for having an affected mother was 6.0 (95% CI 3.0 to 11.9). The risk of developing ovarian cancer by age 75 is about 7.5% for women younger than 45 with an affected mother. The risk declines after this age, but remains substantially higher than the general population.

More than one affected relative

The pooled relative risk from two studies for having more than one affected relative was 11.7 (95% CI 5.3 to 25.9). The risk of developing ovarian cancer by age 75 is about 14% for women younger than 45 with more than one affected relative. The risk declines after this age, but remains substantially higher than the general population.

Mother of affected daughter

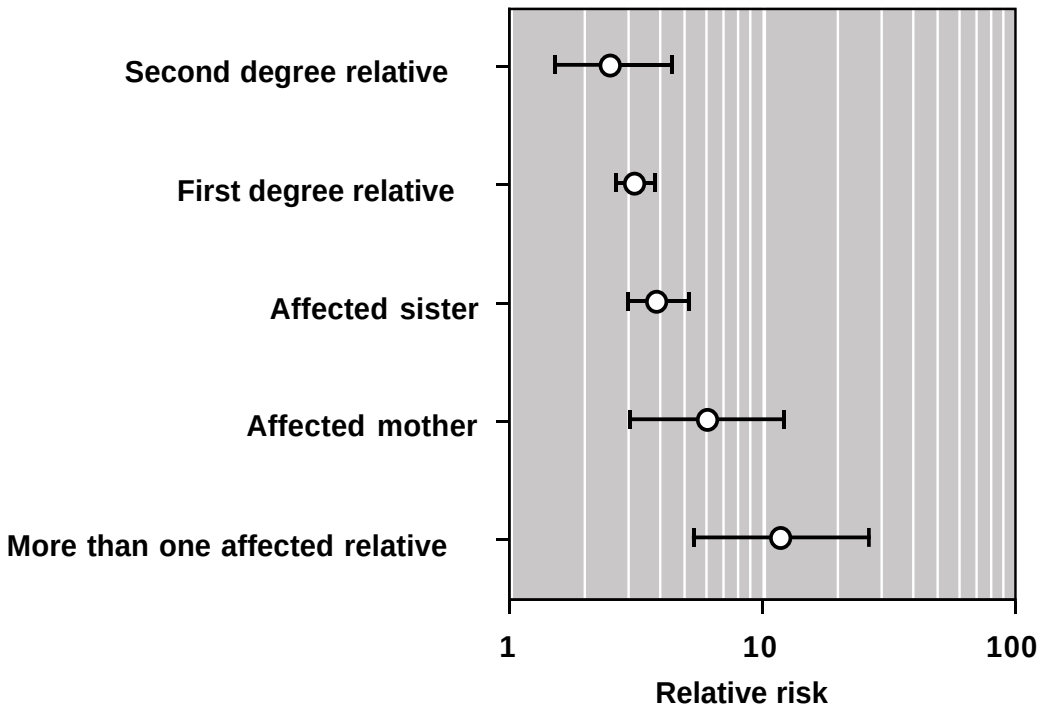
Three studies looked at this. The relative risk was 1.1 (95% CI 0.8 to 1.6)

Comment

This is an interesting and useful paper, and may be helpful, as the authors say, in making decisions about the removal of ovaries at hysterectomy. Many women will find it easy to understand as well, and are likely to appreciate the information the paper contains. It presents risk information well, but ***Bandolier*** cannot immediately reproduce the pictorial methods it uses.

Reference:
1 JF Stratton, P Pharoah, SK Smith, D Easton, B Ponder. A systematic review and meta-analysis of family history and risk of ovarian cancer. British Journal of Obstetrics and Gynaecology 1998 105: 493-9.

Relative risk for ovarian cancer for a woman with relatives diagnosed with ovarian cancer



NUMBER NEEDED TO TREAT OR HARM (NNT/NNH)

A number of new readers have asked for ***Bandolier*** to revisit the calculation of NNTs. As a reminder, ***Bandolier*** gave a full account in issue 36, and that issue is included in the second ***Bandolier*** annual (see below). Many GPs are familiar with NNTs, and a high percentage feel comfortable with understanding NNTs and explaining them to others. A report from Wessex on this is opposite.

Outcomes

Statistical ways to express outcomes of clinical trials include p values, odds ratios, relative risk and relative risk reduction or increase. All may have their place, but they are difficult outputs for the non-specialist to interpret. In order to overcome this, the number needed to treat is increasingly being used. The NNT, as the name implies, is an estimate of the number of patients that would need to be given a treatment for one of them to achieve a desired outcome who would not have achieved it with control. The NNT should specify the characteristics of patients being treated, the intervention and its duration, and the outcome being measured [1].

Calculation

For an analgesic trial, the NNT may be calculated very simply as:

NNT = 1/(the proportion of patients with at least 50% pain relief with analgesic minus the proportion of patients with at least 50% pain relief with placebo)

Taking a hypothetical example from a randomised trial:

- 50 patients were given ibuprofen, and 27 of them had more than 50% pain relief over 6 hours.
- 50 patients were given placebo, and 10 of them had more than 50% pain relief over 6 hours.

The NNT is therefore
$$1/((27/50) - (10/50))$$
$$= 1/(0.54 - 0.20)$$
$$= 1/0.34$$
$$= 2.9$$

The best NNT would, of course, be 1, when every patient with treatment benefited, but no patient given control benefited. Generally NNTs between 2 and 5 are indicative of effective treatments, but NNTs of 20, 50 or 100 may be useful for prophylactic treatments, like interventions to reduce death after heart attack. It all depends on the intervention and the consequences.

Harm

For adverse effects, we can calculate a number needed to harm (NNH), in exactly the same way as an NNT. For an NNH, large numbers are obviously better than small numbers, because that means that the adverse effect occurs with less frequency.

Comment

NNTs and NNHs are a useful common currency for describing results from trials, and especially reviews and meta-analyses. They also allow us to impose our own values by describing what we believe to be a useful outcome.

Reference:
1 HJ McQuay, RA Moore. Using numerical results from systematic reviews in clinical practice. Annals of Internal Medicine 1997 126: 712-720.

SUBSCRIBING TO *BANDOLIER*

There are two different routes - follow route 1 if you work for the NHS and route 2 if you are non-NHS.

Route 1: NHS employees

England and Wales

Bandolier is distributed free within the NHS, but arrangements differ slightly between regions. The ***Bandolier*** office will always try and sort this out for you, but if you live and work in South & West, Trent, North & West or Wales, then contact regional R&D Directorates direct as ***Bandolier*** is sent out locally. For Anglia & Oxford, North & South Thames, West Midlands and Northern & Yorkshire, ***Bandolier*** is able to arrange distribution. All will change with the new Regions.

Scotland and Northern Ireland

In Scotland and Northern Ireland, you have to pay for ***Bandolier*** unless you make arrangements specifically with the appropriate offices. The 1998 cost for single subscriptions is £36 a year, but ***Bandolier*** is charged at cost for bulk distribution to Health Authorities and Trusts.

Route 2: Non-NHS subscriptions

Bandolier is distributed outside the NHS by Hayward Medical Communications, Rosemary House, Lanwades Park, Kentford, Near Newmarket, Suffolk CB8 7PW. Information on subscriptions (1998 price UK £36 a year, Eire and Overseas £72 a year) and payment from:

Jan Lister
email bandolier@hayward.co.uk
tel: 44 1638 751515 fax: 44 1638 751517

Bound Volumes

The second ***Bandolier*** volume of collected issues (21 - 34) is available. It can be obtained by sending a cheque for £14.00, made out to Oxfordshire Health, to: Mrs Eileen Neail, ***Bandolier***, Pain Relief, The Churchill, Headington, Oxford OX3 7LJ, UK. We cannot chase paper, so don't ask us to send invoices; the only way you can get this limited edition collector's item is to send a cheque with your order. Alternatively, to use a credit card or from overseas (£18 overseas) contact Jan Lister at Hayward Medical.

Where do GPs get evidence-based information?

There is a plethora of sources of evidence-based information for doctors. Do they use any? How much do they use? What is the biggest influence? How is information best presented?

These questions, and more, have been answered in a thorough and thoughtful review of Wessex GPs by Alastair McColl and his colleagues from Southampton [1]. Structured questionnaires were sent to a quarter of Wessex GP principals, and 302 (67%) replied. The results obtained were the following:

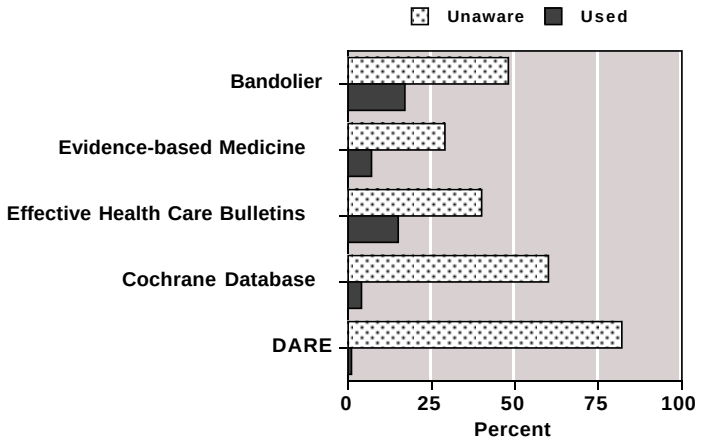
Attitudes

GPs predominately welcomed evidence-based medicine, considered that it improved patient care and that research findings were useful in the day-to-day management of patients. The median value for the estimated percentage of the respondents' clinical practice that was evidence-based was 50%.

Awareness

There was a low level of awareness of evidence-based resources (Figure 1). Over 80% were unaware of DARE or Evidence-based Purchasing (the S&W Region's own newsletter). ***Bandolier***, *Evidence-based Medicine* and *Effective Health Care Bulletins* were the sources with the highest awareness and the highest rating for being used to help in clinical decision-making (Figure 1).

Figure 1: Extremes of awareness of evidence-based resources in 302 Wessex GPs



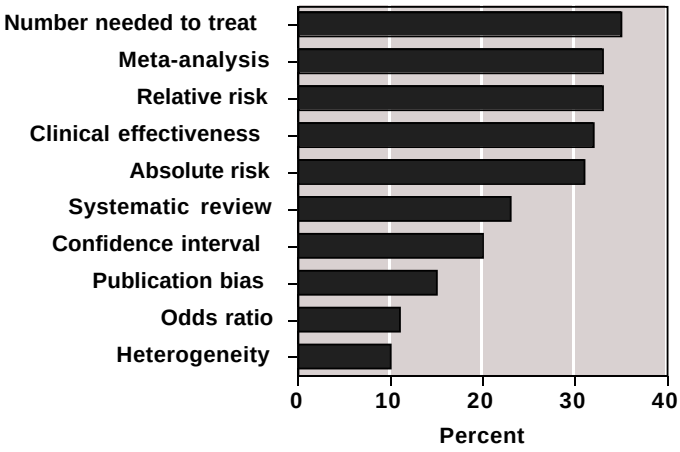
Access

Few GPs had access to searching tools in their surgeries. Only 20% had access to MEDLINE and 17% had access to the Internet. Slightly more than this (21% and 29% respectively) had access at home.

Approach

The questionnaire also asked some penetrating questions about GPs' knowledge of technical terms used in evidence-based medicine (things like odds ratios, heterogeneity and the like). They used a very high hurdle - whether respondents understood the term and could explain it to others. Of all the terms, the one which came out top of this stiff test was the number needed to treat (NNT), with 35% of GPs being able to understand it and explain it to others (Figure 2).

Figure 2: Percentage of 302 GPs in Wessex able to understand and explain technical terms to others



Acquiring more evidence

Few practitioners thought it appropriate for them to identify and appraise primary literature. Most wanted established evidence-based guidelines developed by colleagues, where good evidence and knowledge can be combined with other factors into local wisdom.

Comment

All this makes pretty good sense. General practitioners have been at the forefront of efforts to provide evidence-based education (perhaps reflecting a bias towards prophylaxis - small numbers of events in large populations - in the public health community who have been most enthused by EBM). So finding that, after four years or so, they like it and think it useful is rewarding. That is underpinned by the fact that about a third of these GPs feel able to tackle technical terms head on - not just pretend some understanding. And large numbers of GPs wanted to learn more.

But this survey also gives an idea about the way in which GPs like information presented, which is changing fast. Four years ago the NNT was largely unheard of: now it is the favoured term by Wessex GPs. That is a remarkable change in a short time.

It is clear that what GPs want is evidence presented in short, understandable packets, with recognised quality standards, using understandable descriptions (like NNTs) of useful clinical outcomes, so that evidence can be incorporated easily and quickly into local guidelines and practice.

Reference:

1 A McColl, H Smith, P White, J Field. General practitioners' perceptions of the route to evidence based medicine: a questionnaire survey. BMJ 1998 316:361-5.

MINDSTRETCHER

Most of us can recall the exam question that starts "Compare and contrast...." This month ***Bandolier*** has selected two systematic reviews which make good reading for comparing and contrasting those things which make them good and bad. Below we give ***Bandolier***'s views, but others will have their own opinions, and the two together could make good teaching material.

Dynamic exercise therapy

This review, from Holland, examined dynamic exercise therapy in rheumatoid arthritis [1]. It sought randomised trials about using exercise for at least six weeks through an extensive searching process, and all studies were examined using a 10-point quality score that Dutch reviewers use.

They found six studies, and tell us all about them. In particular they examined the different outcome measures, including both surrogate measures like aerobic capacity, measures of disease activity like pain or ESR, and functional outcomes like walking distance.

The results are disappointing, with some statistical differences with surrogate markers, but no consistent beneficial effects for measures of disease activity or functional ability. What the authors do, however, in discussing the results, is to suggest that the numerous (75) methods of describing outcomes in trials like this needs to be sorted out, so that results may be interpreted better in the future.

Acupuncture in dental pain

Possibly the first systematic review of acupuncture in dental pain [2] again used a comprehensive searching strategy for relevant papers. They were scored using a reliable 5-point scale. Included studies had to be "controlled, conducted in humans and if they tested acupuncture as a treatment of dental pain".

They found 16 trials, of which 11 were randomised and 3 were double blind. Five studies scored zero on the five-point scale, while three scored three points, and none scored four or five points. Detailed tables give information about the 16 trials, their design, the outcomes and the main results for pain and adverse effects.

The conclusion, without pooling data, was that acupuncture can alleviate dental pain, and that future research should concentrate on the best technique and the relative efficacy of acupuncture to conventional treatment.

Compare and contrast

What we have here is two systematic reviews, both well searched, one examining an intervention in a chronic painful condition, and the other an intervention in an acute painful condition. They use different quality scores, but in much the same way. One, on exercise therapy, accepts only randomised studies but includes those where concealment of allocation is not hidden (date of birth, for example). Because the intervention cannot be blinded, this may be less of a problem than in other circumstances.

Both, in effect, use a vote count, noting the number of studies which come up with a statistical significance for some outcome in favour of the intervention. Neither comments on the balance between positive and negative findings for different outcomes, and may bias us because of that. The review on dynamic exercise testing at least gives us all the results on which to make up our own minds.

Validity

The review on dynamic exercise therapy bemoans the lack of consistent criteria on which to base a successful outcome. This is essentially a lack of rules or validity criteria. The review of acupuncture ignores established validity criteria for acute pain measurements.

In acute pain we know that for assays to be valid pain has to be moderate or severe in intensity, that established pain measuring methods must be used for between four and six hours, that trials have to be both randomised and double blind, that pain during a procedure may be different from pain after a procedure, and that experimental pain is not clinical pain.

For acupuncture that leaves us with three randomised, double-blind studies out of the 16. Of those one is of three hours duration only, one is pain during drilling, and one is experimental dental pain. Actually not one of these studies would get into a review of trials with the sort of standards of experimentation that we expect for a new analgesic.

Comment

There are deficiencies in these reviews. We know how important things like randomisation and blinding are for eliminating bias. For instance we reported in ***Bandolier*** 37 on transcutaneous electrical nerve stimulation (TENS) that 17 out of 19 non-random studies showed it worked, while 15 out of 17 randomised studies said it didn't.

Yet both these reviews were inclined to accept information from either non-randomised or inadequately randomised studies. Particularly for subjective measures that is inappropriate. Where there is bias, it is not surprising that it all goes the same way, so vote-counting of biased studies is again inappropriate.

It all comes down to quality. The use of quality standards, including known rules for valid trials, is like using a searchlight on a dark night. These two reviews both fail to use the searchlight, in interesting, but different, ways.

References:

1 CH van den Ende, TP Vliet Vlieland, M Munneke, JM Hazes. Dynamic exercise therapy in rheumatoid arthritis: a systematic review. British Journal of Rheumatology 1998 37: 677-87.

2 E Ernst, MH Pittler. The effectiveness of acupuncture in treating acute dental pain: a systematic review. British Dental Journal 1998 184: 443-7.